

## Consortium 2019 State-by-State Comparisons: Smarter Balanced

D. J. McRae, Ph.D.

10/09/19

[Material in red will be updated when data from Vermont become available]

When two consortiums of states were chosen by the US Department of Education in 2010 to develop new statewide assessment systems, one of the purposes was to generate state-by-state comparable achievement data. Roughly 45 states initially signed up for potential use of consortium tests, but by 2015 when the new tests were ready for their first operational use, only 18 states administered the Smarter Balanced tests and 11 states (plus the District of Columbia) administered the PARCC tests, representing just under 50 percent of total K-12 enrollments across the country. As noted in the 2018 version of this document, although a number of former PARCC states still use public domain test questions developed by PARCC, these states no longer operate as a consortium and do not generate comparable results. So, the 2019 version of this document will not include results from former PARCC states. Michigan uses Smarter Balanced public domain test questions, but does not use the full Smarter Balanced protocol and thus is not included in this report. The data for 2019 includes, then, 11 Smarter Balanced states representing 20 percent of K-12 public school enrollments across the country. Only grade 3-8 results are included; for high school tests, only 7 of the 11 Smarter Balanced states administer Smarter Balanced tests, and only 5 administer high school tests at a common grade level.

The data charts on pages 2 and 3 provide state-by-state results for Smarter Balanced states for spring 2019 testing. The results are expressed as a “percent meeting target” grade-by-grade for English Language Arts and Mathematics, along with average percent across grades. On page 3, the average percent across grades for each state for 2015 thru 2019 are provided, as well as the gain scores for 2015-16 through 2018-19. Notes describing the data in the charts are provided at the bottom of page 3. There may be some differences in test administration or reporting practices across states; however, the comparability of scores is sufficient for general comparisons.

It is fair game to average gain scores for ELA and Math for each state to produce annual overall gain scores for 2016, 2017, 2018, and 2019 results. In the early 2000’s, highly respected educational measurement expert Bob Linn testified before Congress that 3- to 4-percentage point annual gains for statewide testing programs could be characterized as good to very good, and 2-point annual gains were typical. With this background, we can interpret each annual gain score as a letter grade based on a 4.0 grade point average (GPA) metric, with 4.0 being an A, 3.0 being a B, 2.0 being a C, 1.0 being a D, and 0.0 being an F. The annual consortium-wide gain scores for 2016 thru 2019 for Smarter Balanced states are provided numerically and graphically on page 4.

Observations on these Smarter Balanced data are provided on pages 5-6.

Smarter Balanced 2019 State-by-State Comparisons [Percent Level 3 & Above]

		English/Language Arts						
	<u>Grade</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>Ave</u>
1	California	49	50	52	49	51	49	50.0
2	Connecticut	54	55	58	55	56	56	55.7
3	Delaware	51	54	57	52	55	52	53.5
4	Hawaii	53	52	57	53	53	52	53.3
5	Idaho	50	52	57	55	58	54	54.3
6	Montana	48	47	54	51	52	48	50.0
7	Nevada	46	49	52	46	50	48	48.5
8	Oregon	47	43	54	52	55	53	50.7
9	South Dakota	48	50	54	51	54	52	51.5
10	Vermont							
11	Washington	55	57	60	57	61	58	58.0
	Averages	50	51	56	52	55	52	52.6

			<u>Mathematics</u>					
1	California	50	45	38	39	38	37	41.2
2	Connecticut	55	53	47	45	46	44	48.3
3	Delaware	53	51	44	38	41	38	44.2
4	Hawaii	56	49	44	41	38	38	44.3
5	Idaho	53	50	45	43	46	41	46.3
6	Montana	49	45	40	39	42	37	42.0
7	Nevada	48	44	37	34	32	30	37.5
8	Oregon	46	43	38	37	40	38	40.3
9	South Dakota	53	49	40	41	46	44	45.5
10	Vermont							
11	Washington	58	54	48	47	49	46	50.3
	Averages	52	48	42	40	42	39	44.0

# Smarter Balanced 2016-19 Gain Scores

		<u>English/Language Arts</u>								
		2015	2016	2017	2018	2019	15-16	16-17	17-18	18-19
<u>State</u>		ELA	ELA	ELA	ELA	ELA	Gain	Gain	Gain	Gain
1	California	42.3	46.7	46.8	48.8	50.0	+4.4	+0.1	+2.0	+1.2
2	Connecticut	xx*	55.8	54.2	55.2	55.7	xx*	-1.6	+1.0	+0.5
3	Delaware	51.7	54.8	53.8	54.0	53.5	+3.1	-1.0	+0.2	-0.5
4	Hawaii	47.7	50.5	49.2	53.3	53.3	+2.8	-1.3	+4.1	0.0
5	Idaho	49.7	51.8	51.2	52.7	54.3	+2.1	-0.6	+1.5	+1.6
6	Montana	xx*	50.0	49.8	50.5	50.0	xx*	-0.2	+0.7	-0.5
7	Nevada	xx*	48.3	46.2	47.0	48.5	xx*	-2.1	+0.8	+1.5
8	Oregon	53.8	53.3	51.5	52.8	50.7	-0.5	-1.8	+1.3	-2.1
9	South Dakota	47.5	51.2	49.8	53.0	51.5	+3.7	-1.4	+3.2	-1.5
10	Vermont	53.7	56.5	52.5	54.4		+2.8	-4.0	+1.9	
11	Washington	55.5	57.8	57.0	57.8	58.0	+2.2	-0.8	+0.8	+0.2

		<u>Mathematics</u>								
		2015	2016	2017	2018	2019	15-16	16-17	17-18	18-19
<u>State</u>		Math	Math	Math	Math	Math	Gain	Gain	Gain	Gain
1	California	34.2	37.3	38.2	40.0	41.2	+3.3	+0.9	+1.8	+1.2
2	Connecticut	40.3	44.2	45.8	46.8	48.3	+3.9	+1.6	+1.0	+1.5
3	Delaware	40.7	43.7	44.5	44.2	44.2	+3.0	+0.8	-0.3	0.0
4	Hawaii	42.2	43.0	43.2	43.7	44.3	+0.8	+0.2	+0.5.	+0.6
5	Idaho	40.8	43.3	43.3	45.3	46.3	+3.5	0.0	+2.0.	+1.0
6	Montana	xx*	41.0	41.2	41.5	42.0	xx*	+0.2	+0.3	+0.5
7	Nevada	xx*	33.8	33.3	36.5	37.5	xx*	-0.5	+3.2	+1.0
8	Oregon	43.5	42.8	41.8	41.7	40.3	-0.7	-1.0	-0.1	-1.4
9	South Dakota	41.2	44.5	45.8	47.5	45.5	+3.3	+1.3	+1.7	-2.0
10	Vermont	43.2	46.7	44.2	45.0		+3.5	-2.5	+0.8	
11	Washington	49.8	51.5	51.2	51.0	50.3	+1.7	-0.3	-0.2	-0.7

## Notes:

All averages and gains are based on Grade 3-8 data.

\*\*Montana and Nevada participated in Smarter Balanced testing in 2015, but both states experienced technology difficulties that prevented generation of representative scores for the entire state. This circumstance prevents calculation of selected gain scores. Connecticut discontinued the Performance Task for the ELA test in 2016, so for comparability reasons the 15-16 ELA gain score is not recorded.

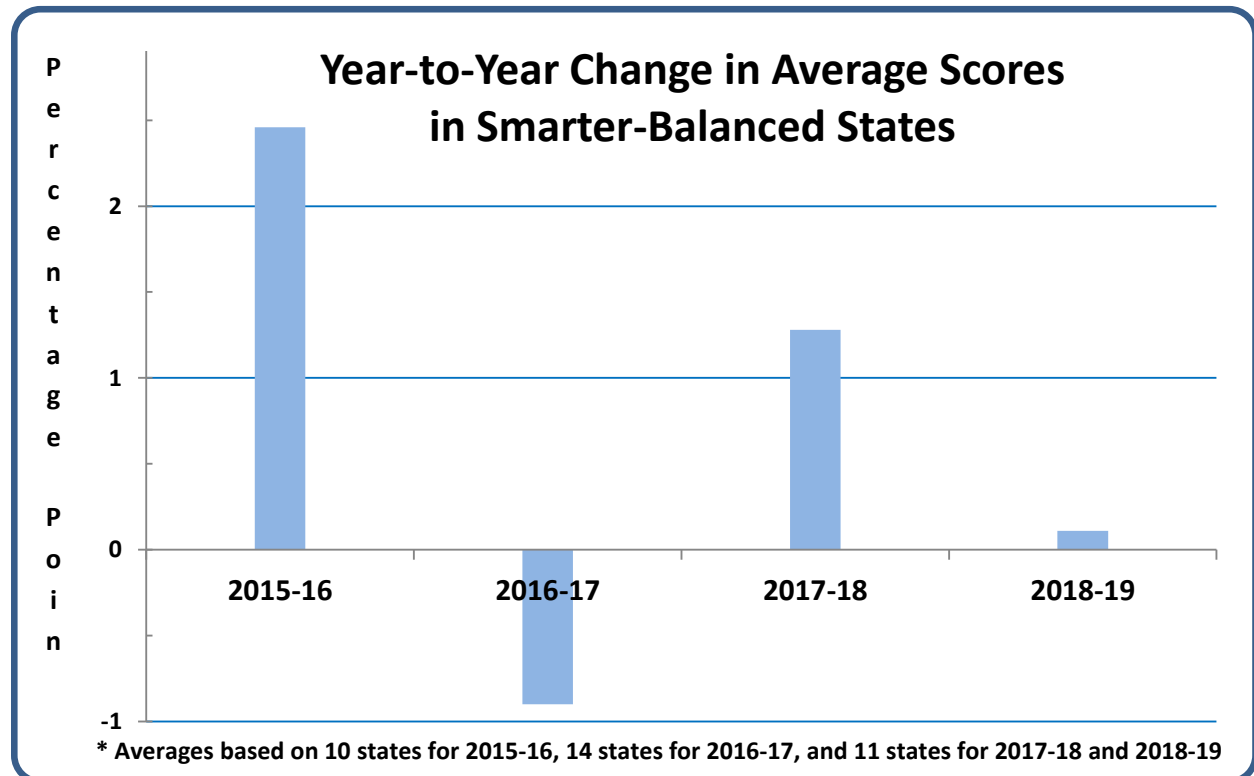
### ELA / Math Gain Score Averages by Year with GPA's and Letter Grades

	Ave Gain 15-16	GPA	Ave Gain 16-17	GPA	Ave Gain 17-18	GPA.	Ave Gain 18-19	GPA
1. California	3.75	A	0.50	D-	1.90	C	+1.20	D
2. Connecticut	xx	--	0.00	F	1.00	D	+1.00	D
3. Delaware	3.05	B	-0.10	F	-0.05	F	-0.25	F
4. Hawaii	1.80	C	-0.55	F	2.30	C+	+0.30	F
5. Idaho	2.30	C+	-0.30	F	1.75	C	+1.30	D+
6. Montana	xx	--	0.00	F	0.50.	D-	+0.00	F
7. Nevada	xx	--	-1.30	F	2.00	C	+1.25	D+
8. Oregon	-0.60	F	-1.40	F	0.60	D-	-1.75	F
9. South Dakota	3.50	A-	-0.05	F	2.45	C+	-1.75	F
10. Vermont	3.65	A-	-3.25	F	1.35	D+		
11. Washington	2.00	C	-0.55	F	0.30	F	+0.25	F
Averages	2.46	C+	-0.90	F	1.28	D+	+0.11	F

#### GPA to Letter Conversions:

A = 3.50 to 4.49, B = 2.50 to 3.49, C = 1.50 to 2.49, D = 0.50 to 1.49, F = Less than 0.50.

Within each range, the higher range of 0.25 to 0.49 merits a plus sign, the lower range of 0.50 to 0.74 merits a minus sign. Equal to or greater than 4.50 merits an A++.



## Observations for Smarter Balanced 2019 State-by-State Comparison Scores

### Smarter Balanced ELA vs Math Results, Trends Across Grades, and Gain Results

Smarter Balanced states averaged 53 percent meeting targets for ELA and 44 percent for Math for the Spring 2019 testing cycle. In 2004, Bob Linn noted differences of 3 to 4 percent are clearly meaningful. Differences of close to 10 percent for ELA and Math clearly meet Linn's criteria, but these differences should primarily be attributed to the threshold scores (cut scores) established for Smarter Balanced tests in 2014 as part of the test development process.

A look at trends across grades for 2019 data shows no obvious trends for ELA results, but do show declining results for Math as the grades increase. These trends across grades are very similar to the trends across grades found for Smarter Balanced states for 2015, 2016, 2017 and 2018 results.

The annual gains for 2016 through 2019 show that consortium-wide Smarter Balanced states had somewhat better than typical gains for 2016, considerably lower gains for 2017 [with most states showing actual declines], recovery for less than typical gains for 2018, and then return to no gain results in 2019. The pattern of gains for Smarter Balanced show extreme changes from year-to-year. The GPA-like metrics translate into letter grades that communicate these differences very accurately.

Due to the number of student scores entering into these consortium-wide calculations, increases or declines in results of perhaps only 0.1 to 0.2 percentage points may be considered "statistically significant." However, the use of theoretical statistical significance calculations for these analyses of statewide test results is questionable. From a practical perspective, increases or decreases of 0.5 percentage points may be considered "meaningful" changes, and increases/decreases of more than 1.0 percentage points should be considered as "very meaningful" changes, similar to typical interpretation of 4-point GPAs.

### Other Considerations

It should be noted that changes in the tests between 2015 and 2019 may substantially affect the gain scores displayed in the charts. The Smarter Balanced submission for federal peer review covering spring 2015 tests "revealed some gaps in item coverage at the low end of the performance spectrum." In January-February 2018, Smarter Balanced released information that the operational item bank used for the spring 2017 testing cycle changed considerably, in an attempt to add easier items to improve coverage at the lower end of the achievement spectrum. Based on Smarter Balanced internal technical data dated October 2016 but not released until January-February 2018, it appeared this effort was not entirely successful. In late March 2018, Smarter Balanced released additional technical information based on analysis of actual spring 2017 item performance. However, this technical information was not consistent with the October 2016 technical information upon which anticipated 2017 item bank performance was based, and the March 2018 technical information did not thoroughly address the differential 2017 Smarter Balanced consortium-wide scores in a way that explained the very large declines in gain scores for 2017. The March 2018 information released by Smarter Balanced as well as reviews of that very technical information by the author are available upon request.

Smarter Balanced has not released information whether the 2019 adaptive item bank and/or adaptive algorithm changed for the 2019 test administration, but the California State Board of Education did post an Information Memo in April 2019 that the 2019 test contained potentially substantive modifications for item types [replacing constructed response test questions that required human scoring with most likely less difficult test questions that could be computer-scored]. Any changes to either the adaptive item bank or adaptive algorithm could inform the interpretation of the annual consortium-wide results on page 4 considerably. In the absence of such information, one can only conclude that the comparability of results for the Smarter Balanced testing system from year-to-year remain suspect for generating accurate year-to-year change data, with extreme declines in gain scores for 2017 particularly noteworthy and suspicious, and the virtually flat 2019 gain scores also suspicious.

Comparison of Smarter Balanced consortium-wide gain scores for 2016 thru 2019 to multi-year gain scores from PARCC states for 2016 thru 2018 as well as the previous California statewide test (STAR) from 2002 thru 2013 provide additional context. Smarter Balanced's 4-year consortium-wide gain scores averaged 0.74 points, letter grade D-. For PARCC, the 2018 version of this data document had 3-year consortium-wide gains that averaged 1.66 percentage points for an overall letter grade of C-. Finally, the author used the same methodology for analyzing California's previous statewide test (STAR) from 2002 thru 2013, and STAR had a 12-year annual gain average of 2.28 percentage points, a letter grade of C+. With differences of +/- 0.50 percentage points indicating clearly meaningful differences, it can be said with great confidence that the stagnant Smarter Balanced consortium-wide gain scores for 2017, 2018, and 2019 do not approach typical two percentage point annual gain criteria articulated by Bob Linn 15 years ago, and are clearly lower than comparable gain scores for the PARCC consortium for 2016 thru 2018 and the 12-year annual gain average for California's previous statewide test for 2002 thru 2013.

Why there has been a 3-year stagnation in Smarter Balanced gain scores is fundamentally unknown. However, from a big picture perspective, it is reasonable to speculate the stagnation may be due to (a) the measurability of Common Core academic content standards targeted by Smarter Balanced tests, (b) less than desirable collective curriculum and instruction implementation for Smarter Balanced states, (c) changes in the test questions in the Smarter Balanced item banks over the years as discussed above, or (d) some combination of all the above.

#### Author Tagline

Doug McRae is a retired educational measurement specialist living in Monterey, California. In his almost 50 years in the K-12 testing field, he has served as an educational testing company executive in charge of the design and development of K-12 tests widely used across the country, as well as an advisor for the design and development of California's STAR statewide testing system which was used from 1998 through 2013. He has a Ph.D. in Quantitative Psychology from the L. L. Thurstone Psychometric Laboratory at the University of North Carolina at Chapel Hill.